JCB Journal of Cell Biology

TOOLS

High-throughput single-cell epigenomic profiling by targeted insertion of promoters (TIP-seq)

Daniel A. Bartlett[1,5], Vishnu Dileep[1], Tetsuya Handa[2], Yasuyuki Ohkawa[3], Hiroshi Kimura[2], Steven Henikoff[4], and David M. Gilbert[1,5]

Chromatin profiling in single cells has been extremely challenging and almost exclusively limited to histone proteins. In cases where single-cell methods have shown promise, many require highly specialized equipment or cell type–specific protocols and are relatively low throughput. Here, we combine the advantages of tagmentation, linear amplification, and combinatorial indexing to produce a high-throughput single-cell DNA binding site mapping method that is simple, inexpensive, and capable of multiplexing several independent samples per experiment. Targeted insertion of promoters sequencing (TIP-seq) uses Tn5 fused to proteinA to insert a T7 RNA polymerase promoter adjacent to a chromatin protein of interest. Linear amplification of flanking DNA with T7 polymerase before sequencing library preparation provides ∼10-fold higher unique reads per single cell compared with other methods. We applied TIP-seq to map histone modifications, RNA polymerase II (RNAPII), and transcription factor CTCF binding sites in single human and mouse cells.

Introduction

The surge in single-cell epigenomics and transcriptomics has been instrumental to our growing understanding of cell fate changes and the nature of cell heterogeneity during development and disease. Unfortunately, tracing the role of chromatin–protein binding in the single-cell context remains uncharted territory due to the lack of efficient and robust methods to map the binding sites of chromatin proteins in single cells. The mounting demand for a highly efficient, robust, and scalable method to map protein binding sites genome-wide in low and single cells is evident in the massive proliferation of low-input chromatin immunoprecipitation (ChIP) methods (Brind'Amour et al., 2015; Cao et al., 2015; van Galen et al., 2016; Grosselin et al., 2019; Zhang et al., 2016), but ChIP methods all suffer from the inherent inefficiency of IP (Baranello et al., 2016; Marinov, 2018). DamID identifies binding sites via transgenic expression of DNA methyltransferase fused to the target protein of interest that methylates nearby adenines. Methylated genomic DNA can be subsequently cut at *DpnII* restriction enzyme sites and the cleaved products sequenced. DamID bypasses the inefficiency of IP but requires cloning, transfection, and in situ expression of the transgenic Dam fusion protein; its resolution is limited by the distribution of *Dpn*II cutting sites; and it requires enzymatic end-preparation and adapter ligation steps before library amplification, leading to sample loss. Methods such as CUT&RUN (Hainer et al., 2019; Skene et al., 2018) and single-cell chromatin immunocleavage sequencing (scChIC-seq; Ku et al., 2019, 2021) tether micrococcal nuclease proteins to the target protein of interest to cleave and solubilize the surrounding chromatin. These approaches also avoid IP but still require library preparation steps that result in sample loss. Tagmentation via Tn5 transposase simultaneously cleaves DNA and attaches Illumina sequencing adapters in a one-step cut-and-paste mechanism. Methods such as CUT&Tag (Kaya-Okur et al., 2019), CoBATCH (Wang et al., 2019), ACT-seq (Carter et al., 2020), and Paired-Tag (Zhu et al., 2021) recruit a proteinA–Tn5 fusion protein (pA-Tn5) to the antibody-bound target protein of interest to produce libraries of fragments that can be subjected to PCR amplification. However, pA-Tn5 inserts forward and reverse adapters in random orientations and at highly variable distances while PCR requires nearby insertions with inverted orientation to amplify efficiently and to prevent PCR biases.

T7 linear amplification has been used for decades to dramatically increase starting material in many contexts where starting material is limited (Eberwine et al., 1992; Van Gelder et al., 1990). A recent chromatin accessibility profiling method demonstrated that modifying the Tn5 transposon to deliver a T7 promoter in place of standard Illumina PCR adapters enabled

[1]Department of Biological Science, Florida State University, Tallahassee, FL; [2]Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan; [3]Division of Transcriptomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; [4]Basic Sciences Division and Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA; [5]San Diego Biomedical Research Institute, La Jolla, CA.

Correspondence to David M. Gilbert: gilbert@sdbri.org; Daniel A. Bartlett: dbart1807@gmail.com

A preprint of this paper was posted to *bioRxiv* on March 19, 2021.

>1,000-fold linear amplification of adjacent genomic DNA (gDNA) by T7 RNA polymerase (RNAP) before reverse transcription and cDNA library preparation (Sos et al., 2016). Linear amplification before PCR library amplification reduced input requirements for profiling DNase I hypersensitive sites from 50,000 down to single cells (Lake et al., 2018). Linear amplification also provides unbiased genomic coverage since only a single ligation/insertion event is required to amplify any given segment (Chen et al., 2017; Rooijers et al., 2019), whereas PCR requires two ligation/insertion events to occur within an optimal distance and in the correct orientation to enable PCR amplification. Linear amplification has been integrated with other methods and shown to significantly increase library DNA yield and detection sensitivity when substituted for conventional amplification (Chen et al., 2017; van Galen et al., 2016; Harada et al., 2019; Hashimshony et al., 2012; Lake et al., 2018; Rooijers et al., 2019). For example, chromatin integration labeling (ChIL) uses a secondary antibody tethered to an oligonucleotide containing a T7 promoter upstream of the Tn5 mosaic binding sequence, and in situ tagmentation inserts the T7 promoter into adjacent DNA. ChIL provides impressive sensitivity and has mapped protein binding sites of select histone modifications in single mouse cells and transcription factor (TF) CTCF in 100 cells (Harada et al., 2019). Unfortunately, ChIL is only compatible with strongly adherent cell culture samples, is not amenable to robotic automation, and requires in-house construction of antibody–oligo conjugates.

We wished to develop a single-cell chromatin profiling method that could be applied to any sample type and scaled to high throughput. CUT&Tag is compatible with any sample type; is readily automatable, fixation independent, inexpensive, and simple to perform; and has shown some promise with single-cell profiling (Kaya-Okur et al., 2019; Carter et al., 2020; Bartosovic et al., 2021; Zhu et al., 2021). Our method, targeted insertion of promoters sequencing (TIP-seq), inserts T7 promoters adjacent to all antibody-bound genomic sites to permit subsequent linear amplification of adjacent gDNA. The use of in vitro transcription (IVT) in TIP-seq leads to substantially increased sensitivity over standard PCR library preparation employed in CUT&Tag. For single-cell mapping, the primary obstacle with all state-of-the-art methods has been sparse genome coverage (Minkina and Shendure, 2019). TIP-seq overcomes this obstacle with a remarkable (10-fold) improvement in library complexity per single cell, achieving high signal-to-noise single-cell data. To increase the throughput of TIP-seq and avoid the expense and equipment dependency of single-cell platforms, we further adapted the method to single-cell combinatorial indexing (sciTIP-seq), which allows high throughput (theoretically up to 153,600 cells per NovaSeq 6000 S4 sequencing run) and the ability to simultaneously profile numerous samples and antibody targets. In summary, sciTIP-seq provides a high-throughput, low-cost, low-background method for single-cell protein mapping with substantial gains in terms of per-cell read coverage.

Results

TIP-seq design

The excellent sensitivity of CUT&Tag is owed to its use of Tn5 to streamline library preparation by directly inserting PCR sequencing adapters in situ, but paradoxically, its sensitivity is inherently limited by PCR. Since Tn5 inserts adapters (mosaic end [ME]-A/B) in random orientations, roughly half the targets do not have adapters in the correct orientation to amplify. Furthermore, PCR library preparation is extremely sensitive to size variations of amplicons: When two adjacent transposition events occur too far apart, they will not amplify efficiently during PCR or sequencing cluster generation, but if too close, they will exponentially bias library coverage due to increased PCR amplification and clustering efficiency of short fragments. Target regions that are small in length (i.e., TF binding sites) also contend with having a decreased likelihood of receiving multiple transposition events to overcome the 50% efficiency due to adapter orientation. Linear amplification has long been used to circumvent amplification and sequence composition bias resulting from PCR amplification and is particularly beneficial when limited starting DNA is available (van Bakel et al., 2008; Eberwine et al., 1992; Hoeijmakers et al., 2011). In fact, numerous chromatin mapping methods have transitioned to linear amplification for increased sensitivity in single cells, including single-cell RNA sequencing (scRNA-seq [a.k.a. CEL-seq]; Hashimshony et al., 2012), ChIL-seq (Harada et al., 2019), MINT-ChIP (van Galen et al., 2016), DamID (Rooijers et al., 2019), ATAC-seq (a.k.a. THS-seq; Lake et al., 2018), and whole-genome sequencing (a.k.a. LIANTI; Chen et al., 2017). Therefore, we hypothesized that adapting CUT&Tag to use linear amplification would greatly improve the sensitivity in single cells. By merging CUT&Tag with existing Tn5 transposition-based linear amplification protocols, we created a custom-designed pA-Tn5 that carries with it a transposon containing a T7 promoter (ME-T7; Fig. 1 B; see Table S1 for custom oligos used in this study adapted from Harada et al., 2019, and Lake et al., 2018), which is recruited to antibody-bound chromatin sites where the Tn5 transposome simultaneously cuts and inserts (tagmentation) the T7 promoter into adjacent gDNA to permit linear amplification (Fig. 1). IVT by T7 RNAP creates roughly 1,000-fold RNA copies of insertion sites, and after reverse transcription, second-strand synthesis, and cDNA fragmentation, TIP-seq amplicons are prepared for sequencing via limited cycle number PCR indexing. With TIP-seq, the distance between two transposition sites is irrelevant since only one T7 promoter insertion is required to amplify the site. We also adapted the pA-Tn5 transposon to deliver only T7 promoter–containing adapters (rather than forward and reverse PCR adapters), thus completely mitigating the efficiency loss of mismatched PCR adapter orientations and distance between insertions. Additionally, linear amplification produces higher fidelity and uniformity since mistakes made during amplification do not themselves become templates to exponentially propagate the mistakes, which translates to higher mappability of single-cell sequencing reads (Chen et al., 2017). Furthermore, TIP-seq is a single-tube protocol to reduce potential sample loss.

Linear amplification substantially increases sensitivity and library coverage over conventional PCR library preparation

To assess the extent to which linear amplification could improve sensitivity over CUT&Tag, we performed TIP-seq on serially

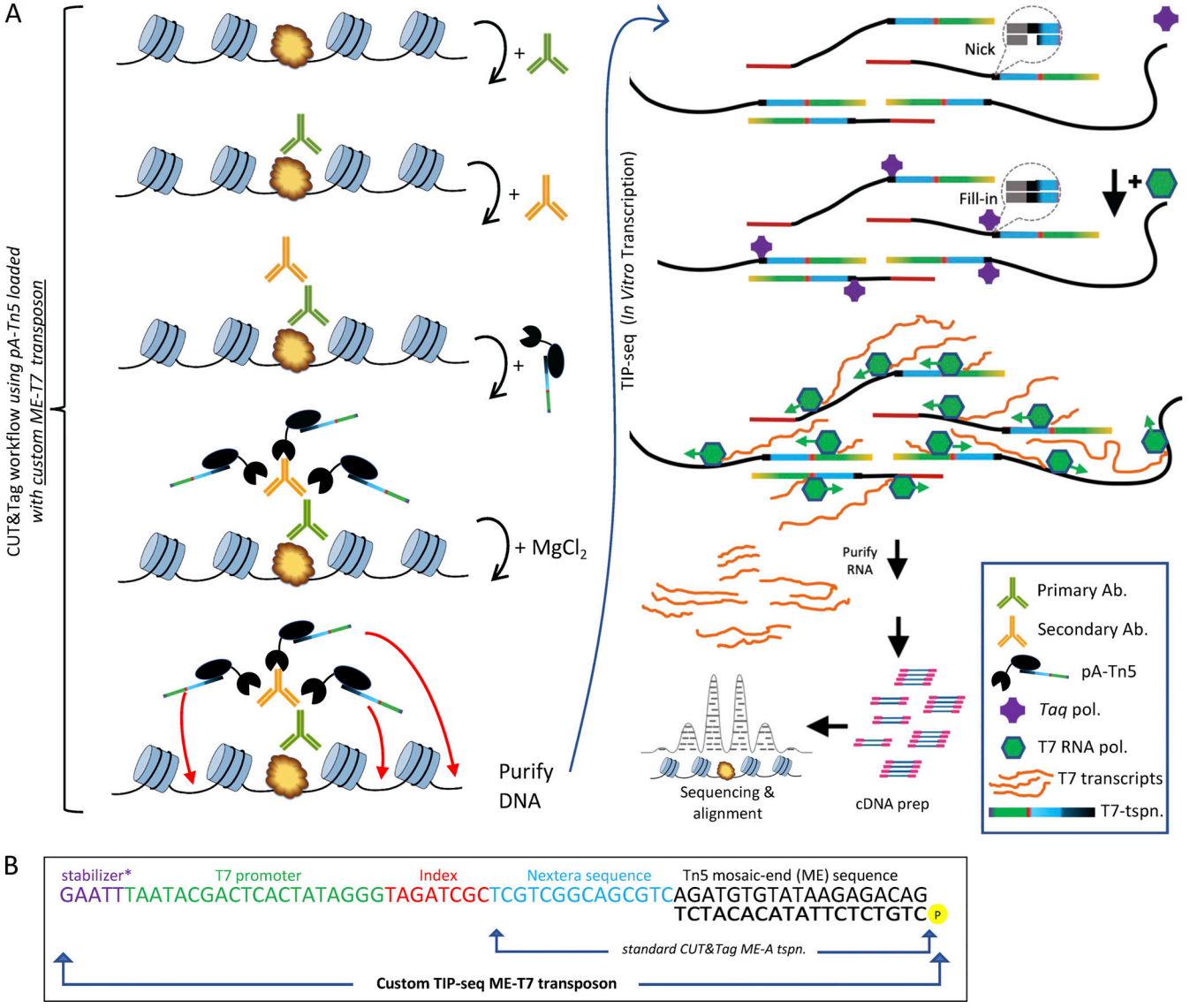


Figure 1. **Overview of TIP-seq, a robust low-cell mapping method that combines CUT&Tag with RNA-mediated linear amplification. (A)** See text for details. **(B)** Custom ME-T7p transposon used in TIP-seq to insert T7 promoters near antibody-bound targets. Standard CUT&Tag ME-A shown for comparison. *Upstream AT-rich stabilizer sequence increases affinity of T7 RNAP for promoter and the efficiency of promoter clearance (Tang et al., 2005). Ab., antibody; pol., polymerase; tspn, transposon.

decreasing numbers (10,000, 1,000, 100, 50) of human HCT116 colorectal cancer cells, targeting the histone modification H3K27me3, and compared it with CUT&Tag using 10,000 HCT116 cells (Fig. 2 A). Remarkably, all cell numbers tested exhibited a higher Pearson correlation to ENCODE ChIP-seq (10 million cells) compared with CUT&Tag from 10,000 cells (Fig. S1 A). We also compared peak numbers and their degree of overlap with ENCODE ChIP-seq peaks and found that all samples overlapped to a similar degree (Fig. S1 B).

To further assess the sensitivity of linear amplification, we performed TIP-seq and CUT&Tag in parallel on serially decreasing numbers of cells targeting the TF CTCF in HCT116. Library preparation and sequencing of CTCF TIP-seq samples produced excellent results for all cell numbers attempted, whereas CUT&Tag libraries failed to yield sufficient library DNA for sequencing with <5,000 cells. Enhanced efficiency was already evident by an ~183-fold increased yield of CTCF library DNA for 5,000-cell TIP-seq vs. 5,000-cell CUT&Tag. Visual inspection of TIP-seq profiles of CTCF showed markedly better agreement with ENCODE ChIP-seq compared with CUT&Tag libraries (Fig. 2 B), with enrichment at a number of foci present in ChIP-seq and TIP-seq being underrepresented or missing entirely from the CUT&Tag sample. This improved performance is explained both by the increased library DNA yield permitting deeper sequencing and, more notably, by the substantially increased library complexity enabled through T7 IVT over PCR amplification. Sequencing reads from CTCF TIP-seq using 5,000 cells yielded ~23.2-fold more unique reads (after filtering for reads originating from duplicated T7 transcripts; i.e., unique Tn5 transposition events) compared with their CUT&Tag

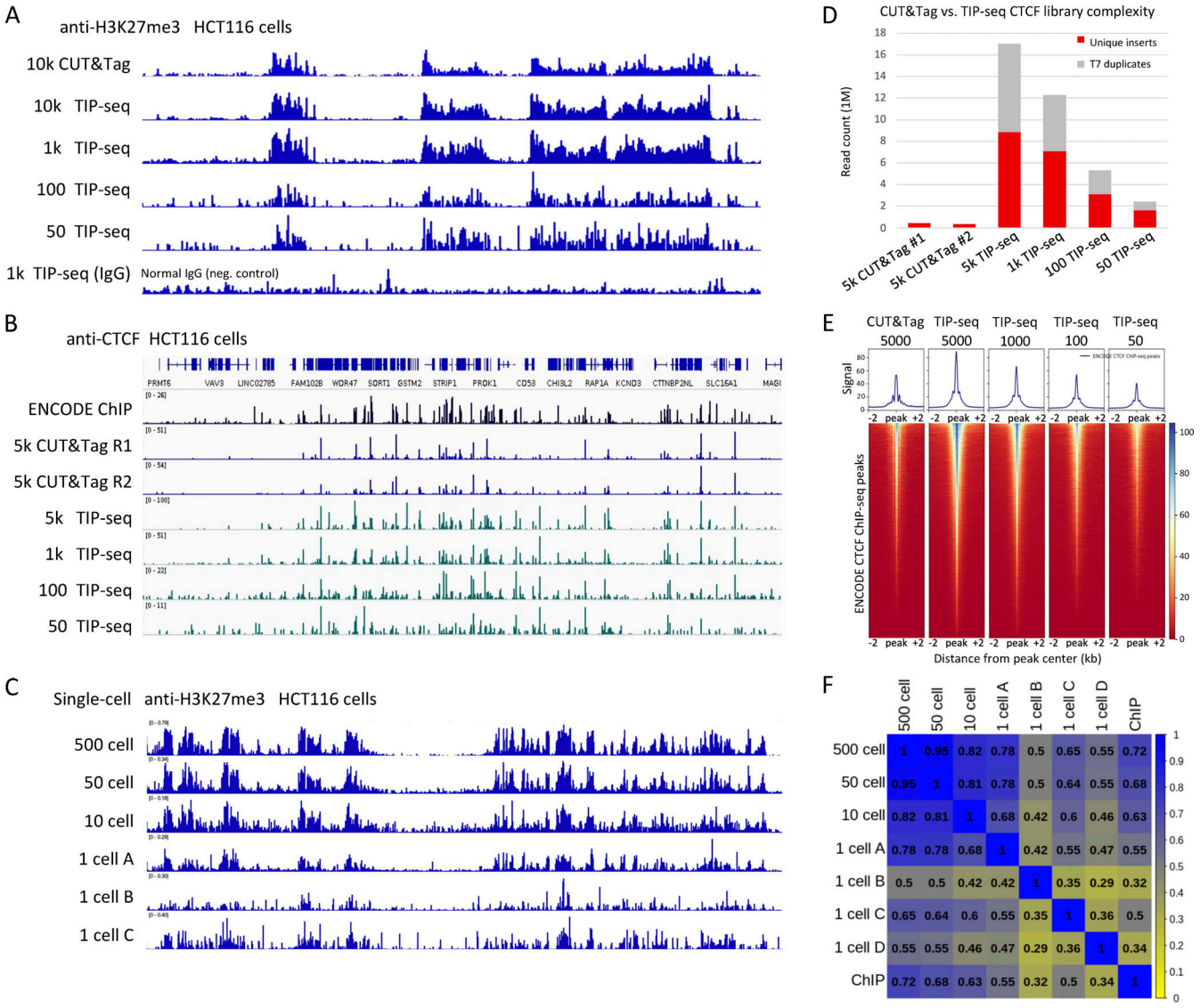


Figure 2. **Bulk TIP-seq substantially increases library complexity and sensitivity over PCR-based library preparation. (A)** Histone modifications. IGV track view across a 3-Mb segment of the human genome for H3K27me3 TIP-seq on 10,000, 1,000, 100, and 50 HCT116 cells. Tracks show CUT&Tag data in 10,000 cells for comparison. TIP-seq for normal IgG in 1,000 cells shown as negative (neg.) control. **(B)** TFs. IGV track view across a 3-Mb segment of the human genome showing TIP-seq targeting CTCF in 5,000, 1,000, 100, and 50 HCT116 cells and two replicates of CUT&Tag data in 5,000 cells performed in parallel with TIP-seq. Top track shows ENCODE ChIP-seq for comparison. CUT&Tag samples from 1,000, 100, and 50 cells failed to yield sufficient library DNA and/or sequencing reads. **(C)** Single cells. IGV track view showing TIP-seq data collected from HCT116 single cells. Cells were processed and tagmented in bulk until cells were FACS sorted into individual PCR tubes for DNA purification and subsequent IVT and library preparation. **(D)** Library complexity of CTCF TIP-seq vs. CUT&Tag as a fraction of unique reads (red) or T7-duplicate reads (gray) over total reads. Samples were processed in parallel, with CUT&Tag (pA-Tn5 loaded with ME-A/B adapters), or TIP-seq (pA-Tn5 loaded with ME-T7 transposons), and after tagmentation and DNA purification, CUT&Tag DNA was PCR amplified using 15 cycles, while TIP-seq DNA was processed as described in Fig. 1 and indexed with nine PCR cycles. Libraries were pooled to equimolar ratios and paired-end sequenced. **(E)** Peak enrichment heatmaps of CTCF TIP-seq vs. CUT&Tag surrounding ±2 kb ENCODE CTCF peaks. Samples were normalized to the sum of per-base read coverage and scaled to 1× genome coverage before plotting heatmaps with deepTools. **(F)** Pearson correlations among bulk TIP-seq, single-cell TIP-seq, and ENCODE ChIP-seq using 50-kb bins.

counterparts (Fig. 2 D), and TIP-seq of CTCF using only 100 and 50 cells received more unique reads than CUT&Tag using 5,000 cells. We then compared the enrichment of reads for the CTCF samples surrounding all CTCF peaks (±2 kb) from bulk ChIP-seq (ENCODE). Compared with the 5,000-cell CUT&Tag, the enrichment of reads at ENCODE CTCF peaks was higher in 5,000-cell and 1,000-cell TIP-seq (~1.8-fold and ~1.3-fold, respectively), and a higher proportion of those ChIP-seq peaks displayed read occupancy (Fig. 2 E). Remarkably, read enrichment surrounding ChIP-seq peaks for 100- and 50-cell TIP-seq resembled that of 5,000-cell CUT&Tag.

All TIP-seq samples profiling CTCF using serially decreasing cell numbers exhibited higher Pearson correlations to ENCODE ChIP-seq (r = 0.65–0.54) compared with CUT&Tag (r = 0.47; Fig. S1 C). We called peaks using SEACR (Meers et al., 2019) and plotted peak numbers and overlap between TIP-seq and

CUT&Tag showing the large majority of peaks overlapping among all samples (Fig. S1 D) but with CUT&Tag having a much lower number of called peaks, as illustrated with browser track profiles in relation to ChIP-seq (Fig. 2 B). We performed a motif search using MEME Suite (Bailey et al., 2009), confirming that the most prevalent motifs from each CTCF library indeed yielded the canonical CTCF binding footprint passing statistical significance (Fig. S1 E).

All efficiency losses and bias accumulation caused by PCR are exacerbated when limited input material is used. Therefore, we next tested how well TIP-seq worked with a single-cell input targeting H3K27me3. We performed TIP-seq in bulk (omitting the use of concanavalin A [conA] magnetic beads) up until termination of the tagmentation reaction to enable FACS sorting of individual tagmented cells into PCR tube strips. Four of five single cells yielded ample DNA to be pooled for sequencing. Again, those four cells delivered profiles remarkably resembling bulk samples (Fig. 2 C) and exhibited Pearson correlations to 500-cell TIP-seq of r = 0.78–0.50 (Fig. 2 F), demonstrating the remarkable power of linear amplification for chromatin profiling single cells.

High-throughput, low-cost sciTIP-seq

A limitation of conventional single-cell preparatory methods is that single cells must be physically separated and compartmentalized before being biochemically processed in separate reaction volumes, causing cost and labor intensity to prohibitively scale linearly with the number of single cells processed. Furthermore, working with single cells, small volumes, and low DNA inputs leads to sample loss. These problems are somewhat mitigated by using specialized microfluidic handlers, but availability and access to expensive and specialized equipment limits the adoption of such techniques. By contrast, sci is a highly scalable method that has been widely adopted to acquire profiles of transcriptomes, genomes, chromatin accessibility, methylomes, and chromosome conformation in tens to hundreds of thousands of single cells without the need for compartmentalization of individual cells (Cao et al., 2017; Cusanovich et al., 2015; Lareau et al., 2019; Mulqueen et al., 2018; Ramani et al., 2017; Vitak et al., 2017; Wang et al., 2019; Weiner et al., 2016; Zhu et al., 2021). Throughput is high, and the cost per cell decreases exponentially with the number of cells processed. sci workflows thus permit affordable and highly scalable throughput and require no specialized equipment that is not widely available, and sci workflows inherently enable sample multiplexing.

We adapted TIP-seq for combinatorial indexing based on existing tagmentation-based sci protocols (Fig. 3; Cusanovich et al., 2015; Lake et al., 2018; Lareau et al., 2019). Briefly, cells were incubated with primary and secondary antibodies in bulk before being split into a 96-well plate (~2,000 cells/well) for the first round of indexing. To add the first index (Index 1), we created 384 uniquely barcoded ME-T7DS transposons (adapted from Lake et al., 2018; T7 promoter optimized based on Tang et al., 2005) such that cells in each well of a 96-well plate received a unique barcode downstream of the T7 promoter. pA-Tn5 tagmentation incorporated the barcoded ME-T7 transposons, then cells were pooled and redistributed randomly via FACS into 96-well plates (15–100 cells/well; up to four 96-well plates can be processed in parallel for index 1 using all 384 ME-T7). DNA was purified and subjected to IVT and cDNA preparation as described for bulk TIP-seq. Index 2 was incorporated via PCR, resulting in library fragments containing both an r5 barcode (Index 1; added via pA-Tn5) and an i5 and i7 barcode (Index 2; added via PCR; Fig. 3 B). The resulting libraries were pooled to equimolar ratios, and sequenced with 29 cycles for index i5 (eight-cycle index run by default) to capture all single-cell index sequences that permit retroactive demultiplexing of reads back to their individual cells of origin.

Multiplexing multiple antigens and single-cell experimental conditions

Since combinatorial indexing begins with the addition of index 1 to cells with a known identity in a 96-well plate, multiple sample types (cells/antibody) can easily be multiplexed at this step and demultiplexed in silico according to which index 1 was added to the corresponding samples in each well. Moreover, by mixing cells from different species, one can accurately assess the cross-contamination (collision) rate at which multiple cells receive the same combination of barcodes. As proof of principle for our multiplexing scheme, we treated five separate pools of mouse F121-9 embryonic stem cells (mESCs) with antibodies against the histone posttranslational modifications H3K27me3, H3K27ac, and H3K9me3; the architectural TF CTCF; and serine 2 phosphorylated RNAPII and a separate pool of human HCT116 cells treated with antibodies against H3K27me3. We also included data from an earlier experiment that included separate pools of HCT116 cells treated with antibodies against H3K27me3, H3K9me3, and H3K27ac using a separate cohort of index combinations that enabled us to pool all these samples together for deep sequencing.

After demultiplexing single-cell FASTQ files based on unique index 1 and index 2 combinations (88% reads assigned to single cells), we obtained data for 5,590 single cells that were aligned to their respective reference genomes using Bowtie 2 (Langmead and Salzberg, 2012), receiving average alignment rates >92%. After filtering reads with low-quality mapping scores and PCR/optical duplicates, all single cells had a mean of 79,606 mapped reads/cell (Fig. S2 A). For an accurate comparison with CUT&Tag datasets, we also filtered out reads originating from duplicated T7 transcripts from the same Tn5 insertion event, since these reads do not provide additional information (albeit, they help to prevent the loss of informative Tn5 insertions that otherwise are lost during CUT&Tag PCR library preparation). Average PCR and T7 duplication rates among all cells were 9.8% and 18.5%, respectively, for an overall 25.3% duplication rate among all cells (Fig. S2, B–G). We then filtered out cells with <1,000 reads (retaining 64% of cells) to obtain 3,557 cells with an overall mean of 38,054 reads/cell (Fig. 4 A). Histone marks, TF CTCF, and RNAPII, respectively, received 50,931, 2,871, and 4,717 mean reads/cell that were used for further analysis (Fig. 4 B). Average PCR and T7 duplication rates after filtering low-read cells were 14.3% and 27.7%, respectively (37.3% overall duplication rate; Fig. S2, C–G), with most cells passing the filter being sequenced near saturation (Fig. S2 F). The high-alignment ratios, high-mapping

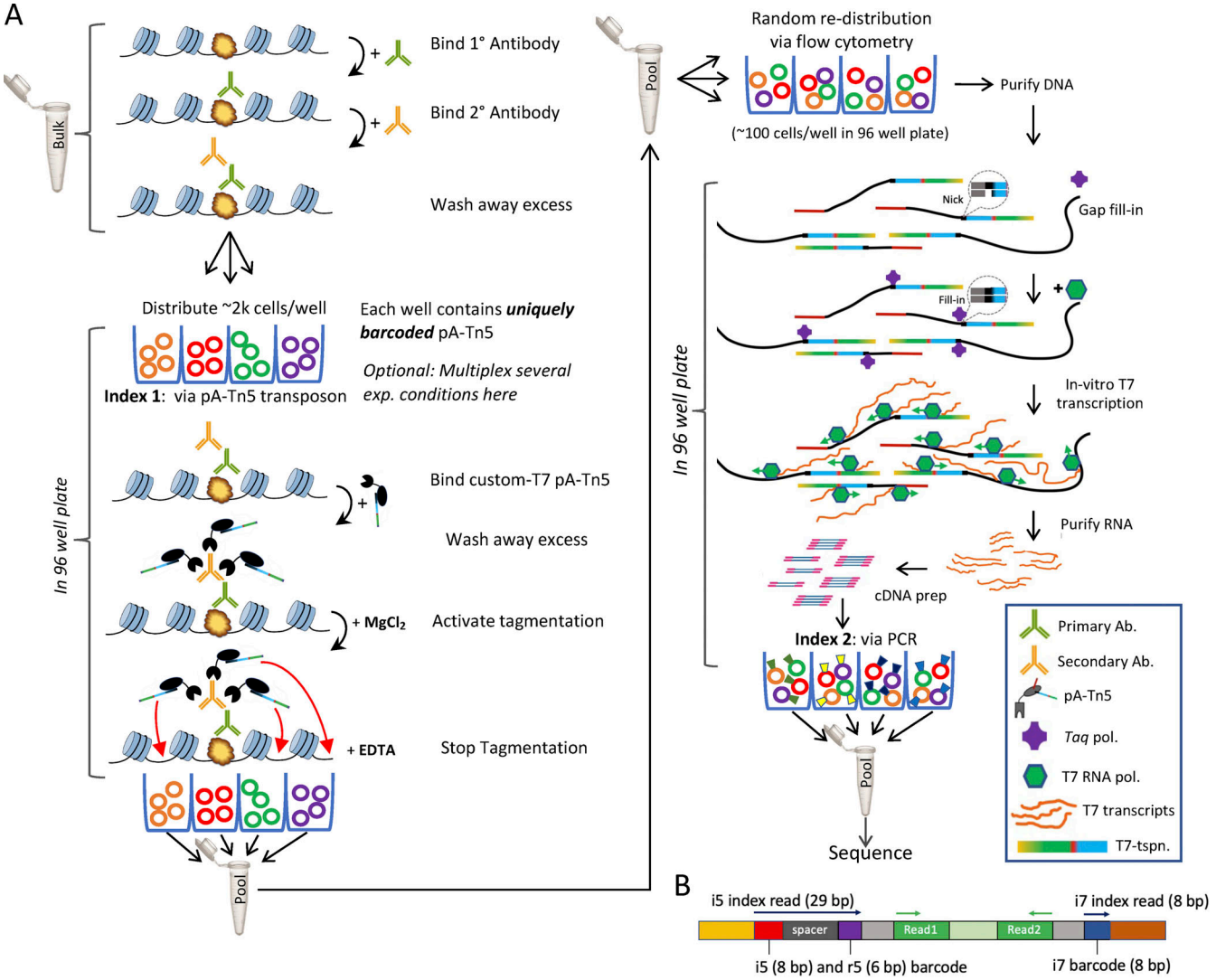


Figure 3. **Overview of sciTIP-seq. (A)** Cells are harvested, permeabilized, and treated with primary and secondary antibodies in bulk. Cells are counted and distributed to a 96-well plate (~2,000 cells/well) where they are incubated with custom, uniquely indexed pA-Tn5. Cells are washed to remove unbound pA-Tn5 before activating tagmentation. Tagmentation is terminated by addition of EDTA, and cells are pooled together and redistributed at random to a new 96-well plate (~15–100 cells/well, depending on number of barcode combinations used during index 1). DNA undergoes a gap-fill reaction via Taq polymerase and IVT via T7 RNAP. RNA is purified and reverse transcribed using a random hexamer primer, then primed for second-strand synthesis using primer complementary to the ME-T7 transposon. ME-B–only Tn5 was used to simultaneously fragment and adapter-tag 3′ end of cDNA to prepare for PCR indexing. **(B)** Resulting library fragments contain an r5 index added during targeted pA-Tn5 tagmentation and an i5 and i7 index added during PCR to enable retroactive demultiplexing of single cell reads. Ab., antibody; exp., experimental; pol., polymerase; tspn, transposon.

quality, and excellent library complexity reflect vastly improved read coverage obtained for single-cell protein mapping data.

We ascertained cross-contamination rates (i.e., collisions) by performing a barnyard analysis on our mixed-species experiment. Collision rates were predicted based on the fraction of reads from individual cells that cross map to either the mouse or the human reference genome. We classified any cell with >10% of its reads aligning to both mouse and human reference genomes as being a collision event. We found a 5.6% collision rate after filtering out cells with <200 reads, indicating low cross-contamination (Fig. S2 H).

Next, we visually inspected how the single-cell data agreed with bulk profiles by first aggregating the reads from all the individual cells, creating a "pseudo-bulk" sample. The pseudo-bulk samples displayed similar profiling to those generated from bulk TIP-seq/ChIP/CUT&Tag data, with the reads from individual cells overwhelmingly falling within enriched regions of bulk samples (Fig. 4 D). We then assessed signal-to-noise ratios of individual cells by calling peaks on the aggregate sample using MACS2 (Zhang et al., 2008) and measured the fraction of unique reads per cell that fell within called peaks (FRiP). Median FRiP scores ranged from 46% to 82%, depending on the sample (60.1% overall; Fig. 4 C), indicative of high signal-to-noise ratios. Although we observed a bimodal distribution of reads per cell in all samples (Fig. S2 A), a few samples retained their cohort of low-read cells after filtering out cells with <1,000 reads (Fig. 2 B). Still, cells in the low-read-count cohort retained high FRiP scores (Fig. S2 I), indicating that they are still of high quality but with reduced coverage more akin to CUT&Tag.

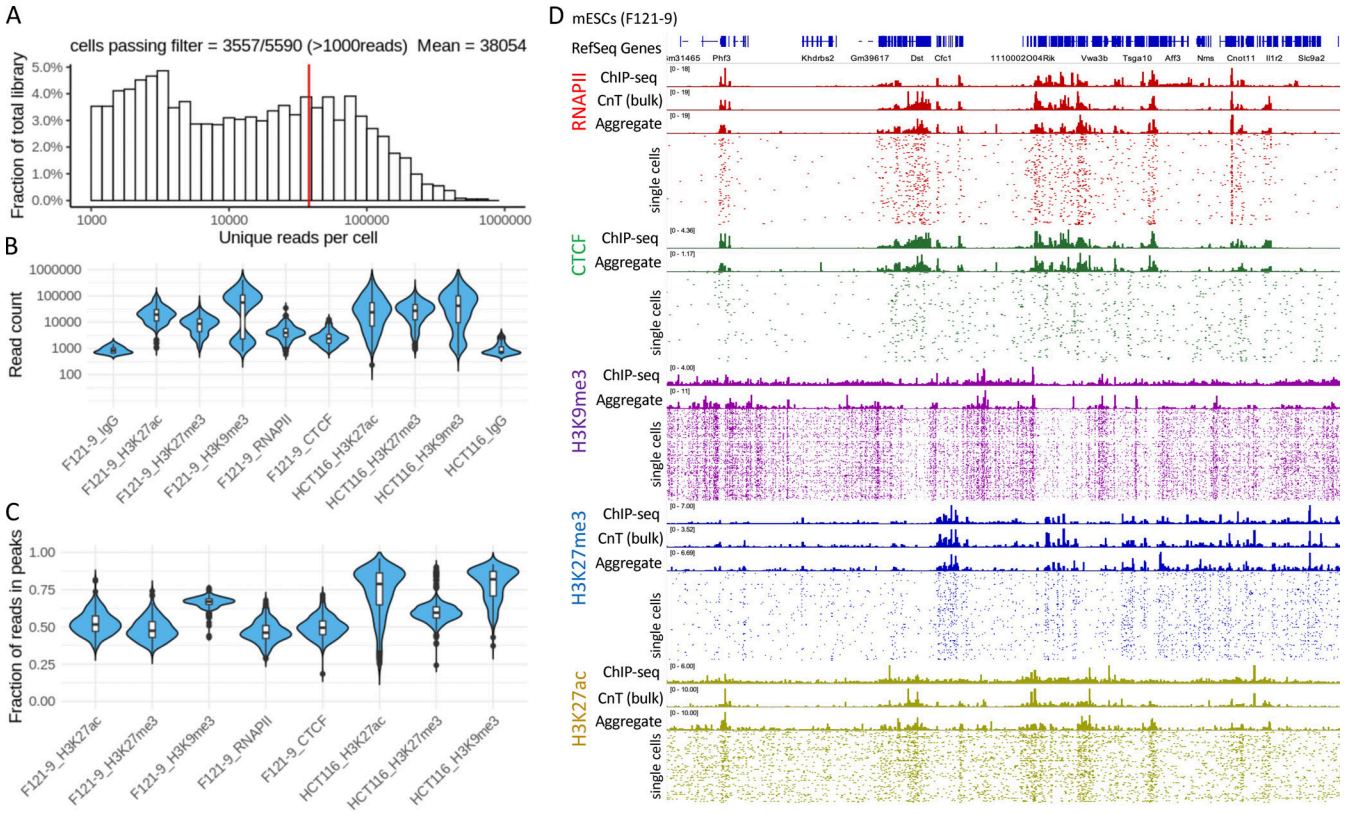


Figure 4. **Multiplexed sciTIP-seq in F121-9 mESCs. (A)** Distribution of unique reads of 3,557 single cells in F121-9 mESCs and human HCT116 cells after removal of PCR/T7 duplicates and filtering out cells with <1,000 reads. Red line indicates the mean (38,054) unique reads per cell. **(B)** Violin plots showing the number of unique reads per cell for each sample after filtering. **(C)** Violin plots showing the FRiP for each respective sample. **(D)** IGV track view across an 11-Mb segment of the mouse genome for sciTIP-seq. Tracks show 100 single cells together with both pseudo-bulk (aggregate) and bulk TIP-seq/CUT&Tag (CnT)/ChIP data (when available) at the representative loci for each sample (RNAPII, CTCF, H3K9me3 H3K27me3, H3K27ac). Bulk ChIP data are from ENCODE and the 4D Nucleome consortium.

To further validate the specificity of sciTIP-seq targeting TF CTCF, we performed a de novo motif search within called peaks using MEME Suite (Bailey et al., 2009) in the pseudo-bulk CTCF sample and a number of single-cell CTCF datasets chosen at random. In both cases, the canonical CTCF binding footprint was identified as the most prevalent motif exceeding statistical significance (Fig. S1 E), demonstrating both specificity and ability to detect motifs in single cells.

We assessed how RNAPII mapping compared with sciTIP-seq and the degree to which it detects heterogeneity in the population. RNAPII is a direct map of transcription, but RNA-seq is a heavily processed, edited, and reverse-transcribed product of transcription, so naturally, there will be differences that have nothing to do with transcriptional heterogeneity. Consistently, uniform manifold approximation and projections (UMAPs) on RNAPII sciTIP-seq showed less heterogeneity than scRNA-seq done in mESCs (Fig. S3 E).

Single-chromosome mapping enabled by high-coverage sciTIP-seq data

In the majority of cases, single-cell data are the average of two chromosomes. The only way to overcome this obstacle is to either use haploid cells, for which there are very few cell lines, or to parse the single nucleotide polymorphisms (SNPs) that exist between maternal and paternal genomes in cases where they have been haplotype phased. Single-chromosome mapping of diploid cells requires much greater read depth to overcome the loss of reads that do not overlap SNPs. The extremely high read coverage of TIP-seq suggested that we might be able to map protein binding sites on single chromosomes. As proof of principle, we parsed the data from F121-9 (which harbors SNPs at a density of one per ~150 bp on average) that represent the polymorphisms between subspecies *Mus musculus* (129) and *Mus castaneus* (Cas). To parse alleles in RNAPII sciTIP data from hybrid F121-9 mESCs, we filtered for reads containing SNPs using the SNPsplit algorithm (Krueger and Andrews, 2016). Reads were aligned to their respective phased genomes and subsequently processed as described in nonparsed sciTIP-seq. Indeed, sites with substantial differences in occupancy between the Cas and 129 alleles were easily identified by visual inspection and confirmed by comparing to parsed bulk CUT&Tag data (Fig. 5).

Comparisons of sciTIP-seq with other state-of-the-art methods

Platforms that automate single-cell partitioning and barcoding have become popular in recent years; however, these require access to specialized equipment and expensive commercial kits unavailable to most researchers. Nonetheless, a recent study (Wu et al., 2021) adapted single-cell

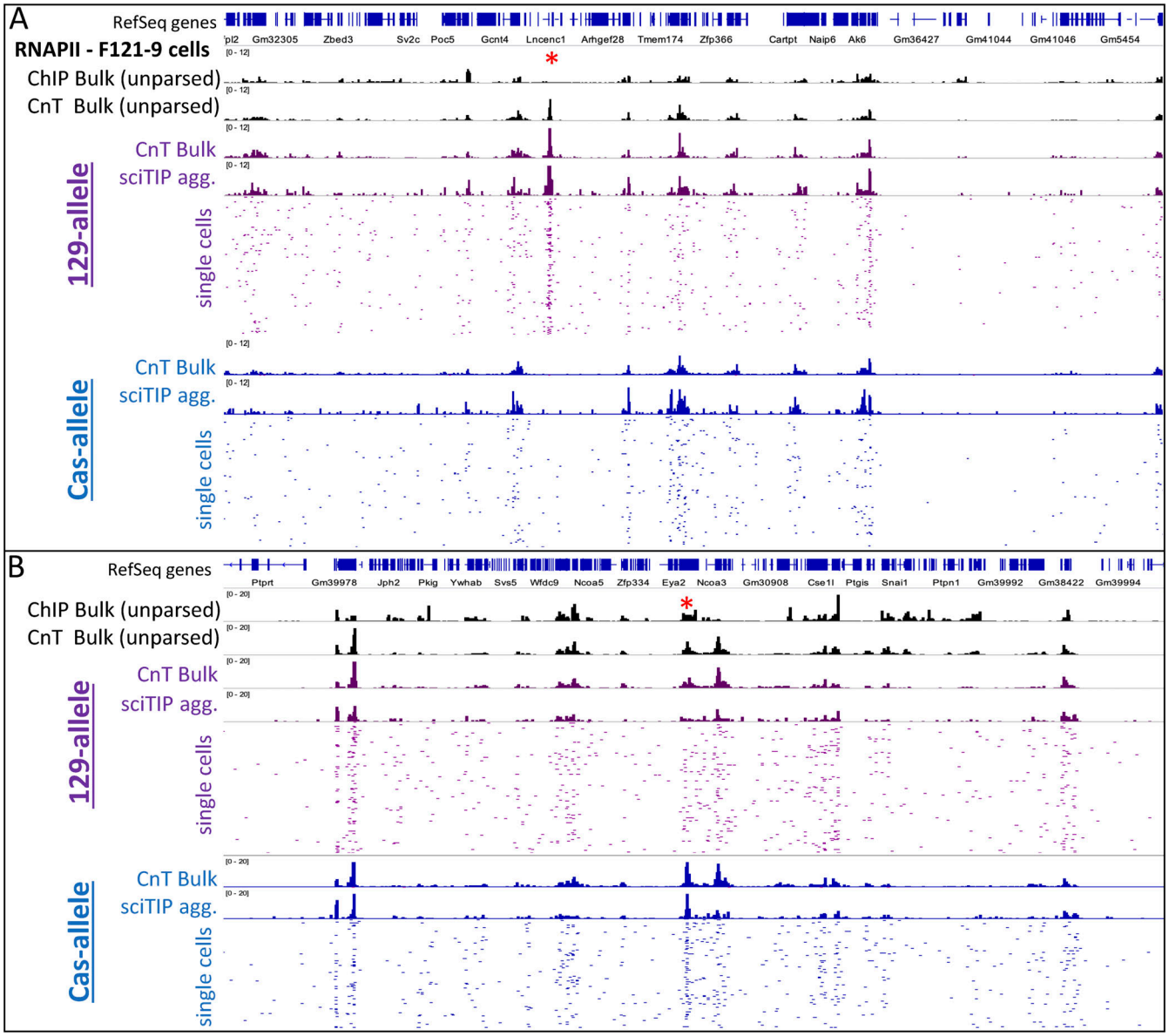


Figure 5. **Single-chromosome mapping of RNAPII to parsed alleles in hybrid F121-9 mESCs. (A)** IGV track view across a 9-Mb segment of the mouse chromosome 13 comparing parsed vs. unparsed bulk CUT&Tag, sciTIP pseudo-bulk, and single cells. Tracks show unparsed bulk ChIP (4D nucleome consortium), unparsed bulk CUT&Tag, parsed bulk CUT&Tag, parsed sciTIP pseudo-bulk (aggregate), and 100 parsed single cells. Red asterisk marks loci with differential RNAPII binding. **(B)** IGV track view across a 7-Mb segment of the mouse chromosome 2 with all else the same as in A. agg., aggregate; CnT, CUT&Tag.

CUT&Tag to use the droplet-based 10x Genomics single-cell ATAC-seq platform. The researchers obtained single-cell CUT&Tag data for H3K27me3 in 9,917 mixed peripheral blood mononuclear cells with 1,110 unique reads/cell, whereas sciTIP-seq yielded 10-fold higher coverage at 11,144 unique reads/cell for H3K27me3. A second study (Bartosovic et al., 2021) adapted scCUT&Tag for the 10x Genomics platform targeting various histone modifications (H3K4me3, H3K27ac, H3K36me3, H3K27me3) with <450 unique reads/cell, 58-fold lower than TIP-seq, with a median of 26,136 unique reads/cell for TIP-seq on histone modifications. These researchers also targeted TFs Olig2 and cohesion complex component Rad21, obtaining 48 and 240 unique reads/cell, respectively, compared with TIP-seq targeting TF CTCF and RNAPII and yielding 2,871 and 4,717 unique reads/cell, respectively.

CUT&Tag has also been adapted for combinatorial indexing in three recent studies (ACT-seq, Carter et al., 2020; CoBATCH, Wang et al., 2019; Paired-Tag, Zhu et al., 2021). Unlike sciTIP-seq, these methods use PCR rather than IVT amplification and yield lower library coverage. Carter et al., 2020 reported 2,500 unique reads/cell (85% duplication rate) for 1,246 cells, whereas sciTIP-seq yields 25-fold more unique reads/cell. Wang et al. (2019) reported between 7,500 and 12,000 unique reads/cell (48% duplication rate) for 2,161 cells (albeit this is an overestimation since cells with <3,000 reads were removed from this statistic).

Paired-Tag reported the highest complexity single-cell CUT&Tag data published to date (Zhu et al., 2021), so we downloaded Paired-Tag datasets from mouse brain tissue samples to perform a comparison with sciTIP-seq on shared epitopes. We filtered cells from both libraries based on what the Paired-Tag study used and calculated FRiP with identical parameters between library types. TIP-seq yielded ∼10-fold more unique reads per cell (64% duplication rate for Paired-Tag; Fig. S3, A and C) and FRiP scores on par or exceeding Paired-Tag (Fig. S3, B and D). One caveat is that Paired-Tag was performed on brain tissue samples compared with TIP-seq with cultured cells. Indeed, a rigorous comparison between methods should be for the same cell type and the same monoclonal antibody.

In summary, while there are differences in epitopes and sample types that render these comparisons indirect, TIP-seq consistently provides much higher complexity libraries for single-cell data.

Discussion

We show that TIP-seq produces single-cell sequencing data for both histones and TFs with dramatically higher read coverage per cell, substantially higher mappability and library complexity, and considerably lower background compared with all current chromatin mapping methods. Other single-cell methods traditionally require merging the reads of hundreds of single cells to acquire an aggregate profile that resembles bulk profiles. Remarkably, TIP-seq yields single-cell profiles resembling those of bulk samples with just a single sufficiently sequenced cell due to the increased read coverage per cell. It is likely that TIP-seq can achieve the coverage necessary to perform studies of single-cell variation in allele-specific epigenomic features, such as both imprinted and random monoallelic expression. Multiple antigens can be mapped simultaneously in the same single cell by preloading primary antibodies with uniquely barcoded pA-Tn5 (Gopalan et al., 2021 *Preprint*), or covalently binding barcoded ME adapters to specific primary antibodies (Meers et al., 2021 *Preprint*). Finally, TIP-seq will also be adaptable for other platforms, such as 10x Genomics, and can be adapted to capture HiC or RNA-seq and sciTIP protein mapping data from the same single cells.

Materials and methods

pA-Tn5 production

Our pA-Tn5 (ME unloaded) was provided by Steve Henikoff (Fred Hutchinson Cancer Research Center, Seattle, WA) and prepared as described in Kaya-Okur et al. (2019).

Preparation of barcoded T7 promoter pA-Tn5 transposons

The 384 uniquely barcoded transposome complexes were adapted from Lake et al. (2018) and assembled according to Kaya-Okur et al. (2019). Briefly, each unique ME T7 promoter adapter oligonucleotides (ME-T7; see Table S1) were annealed to the ME reverse oligonucleotide (ME-rev, 5′-[phos]CTGTCTCTTATACACATCT-3′; see Table S1). For annealing, ME-T7 and ME-rev oligos were diluted to 200 µM in annealing buffer (10 mM Tris, pH 8, 50 mM NaCl, 1 mM EDTA) and mixed 1:1, resulting in 100 µM annealed product. Annealing was performed by denaturing oligo mixture for 5 min at 95°C and slowly cooling to 20°C using a ramp rate of 0.1°C/s. Annealed transposons were aliquoted and stored long term at −20°C.

pA-Tn5–adapter complex formation

For bulk TIP-seq, transposome complex formation was prepared in large batches by adding 16 µl of 100 µl preannealed ME-T7 oligonucleotides with 100 µl of 5.5 µM naked pA-Tn5 fusion protein, stored at −20°C, and used over a period of several months. For sciTIP-seq, transposome complex formation was prepared within 48 h of the experiment by adding 0.5 µl of 100 µM preannealed ME-T7 transposon to 0.5 µl naked 5.5 µM pA-Tn5 in a multiwell plate and incubating for 1 h at room temperature at 300 rpm. Plates were stored on ice until use.

CUT&Tag

CUT&Tag was performed as described in Kaya-Okur et al. (2019). The detailed step-by-step protocol is available on protocols.io (https://www.protocols.io/view/bench-top-cut-amp-tag-bcuhiwt6). Briefly, cell cultures were harvested and rinsed once with an equal volume of room-temperature PBS. Cells were moved to a 2-ml tube and washed twice with room-temperature wash buffer (20 mM Hepes, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1× protease inhibitor cocktail) before counting with a hemocytometer and aliquoting desired cell numbers into fresh 1.5-ml LoBind Eppendorf tubes containing 1 ml room-temperature wash buffer. ConA-coated magnetic beads (BP531; Bangs Laboratories) were prepared as previously described (Skene and Henikoff, 2017), and 10 µl of washed bead slurry was added to aliquots of washed cells and rotated for 10 min at room temperature to allow beads to bind to cells. Thereafter, solution changes and washes were performed by gently pulse spinning (<100 x*g*) tubes and placing them in a magnetic rack for 3 min to collect cells on the side of the tube wall before removing supernatant. To resuspend, a small amount of liquid (∼30 µl) was added to the side of the tube to let it run over the bead clump followed by gentle tapping of the tube until the clump was mostly broken up and resuspended before adding the remainder of liquid to the tube. Bead-bound cells were incubated with primary antibody (1:100) in 100 µl antibody buffer (wash buffer + 0.01% digitonin [Dig], 2 mM EDTA, and 1% BSA) overnight at 4°C with rotation. Antibody buffer was removed and replaced with 100 µl Dig-wash buffer (wash buffer + 0.01% Dig) containing guinea pig anti-rabbit IgG secondary antibody (1:100) and rotated for 1 h at room temperature. Cells were washed twice in 1 ml Dig-wash buffer and then resuspended in 100 µl Dig-300 buffer (0.01% Dig, 20 mM Hepes, pH 7.5, 300 mM NaCl, 0.5 mM spermidine, 1× protease inhibitor cocktail) containing pA-Tn5 (loaded with ME-A/B adapters; 1:100) and incubated at room temperature with rotation for 1 h to allow pA-Tn5 to bind to antibody-bound sites. Cells were washed three times with 1 ml Dig-300 buffer to remove unbound pA-Tn5. Tagmentation was activated by resuspension in 100 µl Tag buffer (Dig-300 buffer + 10 mM $MgCl_2$) for 1 h at 37°C and was terminated by adding 3.3 µl of

0.5 M EDTA. Cells were mixed with 2 µl of 10% SDS (0.2% final) and 0.84 µl of 20 mg/ml proteinase K and incubated at 50°C for 30 min or overnight at 37°C before being purified using 2.0× volume solid phase reversible immobilization (SPRI) beads (A63881; Beckman Coulter). Libraries were amplified using 21 µl CUT&Tag DNA and adding 2 µl of uniquely barcoded i5 and i7 primers and 25 µl of NEBNext High-Fidelity 2X PCR Master Mix (M0541L; New England Biolabs). The samples were PCR amplified using the following thermocycler program: 72°C for 5 min (gap filling), 98°C for 30 s, 14 cycles at 98°C for 10 s and 63°C for 30 s, final extension at 72°C for 1 min, and hold at 8°C (12–15 cycles). Amplified libraries were cleaned up with a 1.1× volume of SPRI beads, washed 2× with 80% EtOH, and eluted into 30 µl 10 mM Tris (pH 8.0).

Bulk TIP-seq

All steps before DNA purification of tagmented cells are the same as in CUT&Tag, with the only exception being the use of a custom pA-Tn5 transposon that contains within it a T7 promoter upstream of the standard Nextera ME-A/B transposons. Refer to the CUT&Tag section for all steps before DNA purification. After SPRI purification of tagmented gDNA, the DNA and beads were resuspended in 8 µl water (SPRI beads remain in solution for IVT and cDNA prep) and gap filled by adding 2 µl of Taq 5X Master Mix (M0285; New England Biolabs) and incubating at 72°C for 3 min. IVT was performed using the HiScribe T7 high-yield synthesis kit (E2040S; New England Biolabs) by adding 2 µl of 100 mM NTP set, 2 µl 10× T7 reaction buffer, 2 µl T7 polymerase mix, and 0.3 µl RNase inhibitor and incubating at 37°C for 16–19 h. RNA was purified by adding 2.0× volume SPRI binding buffer (20% polyethylene glycol 8000, 2.5 M NaCl, 10 mM Tris-HCl, 1 mM EDTA) to reactivate the beads' binding capacity. After washing twice with 80% EtOH and removing all liquid, residual EtOH was dried from RNA and SPRI beads for 3 min before being resuspended in 9 µl RNase-free water. First-strand synthesis was primed by adding 2.5 µl of 20 µM random hexamer to the sample and incubating at 70°C for 3 min, then immediately placing on ice. First-strand synthesis was performed using SMART MMLV Reverse Transcriptase kit (639524; Takara) by adding 4 µl of 5× first-strand synthesis buffer, 2 µl of 10 mM dNTP mix, 2 µl of 100 mM DTT, and 0.5 µl SMART MMLV Reverse Transcriptase. Samples were mixed well and incubated at 22°C for 10 min, 42°C for 60 min, and then terminated at 70°C for 10 min. 1 µl of a 1:10 dilution of 5 U/µl RNase H was added and incubated at 37°C for 20 min to degrade RNA in cDNA–RNA hybrids. Second-strand synthesis was primed by adding 2.5 µl of 20 µM sss_scnXTv2 oligo (anneals to transposon directly downstream of T7 promoter transcription start site) and heating the sample to 65°C for 2 min and placing immediately on ice. Second-strand synthesis was performed by adding 5.9 µl of Taq 5X Master Mix (M0285; New England Biolabs) and incubating at 72°C for 8 min, then cooling on ice. cDNA was purified by the addition of 2.0× SPRI binding buffer, washed twice in 80% EtOH, and resuspended in 7 µl water. Fragmentation and 3′-end adapter tagging of cDNA were performed by adding 2 µl TAPS buffer and 2 µl of 0.7 µM Tn5 (loaded with ME-B adapters only), incubated at 55°C for 6 min, then briefly cooled on ice before adding GuHCl (4 M final concentration) and vortexing to degrade Tn5. DNA was purified again by adding 2.0× volume SPRI binding buffer, washing twice with 80% EtOH, and eluting DNA off the beads into 16 µl water. Eluent was moved to a fresh tube, this time leaving behind SPRI beads to discard. PCR reactions were prepared by adding 20 µl NEBNext High-Fidelity 2X PCR Master Mix (M0541L; New England Biolabs), 2 µl of 10 µM standard Nextera index primers, and 2 µl of 10 µM standard i7 Nextera indexes, for a total volume of 40 µl. The optimal number of PCR cycles was determined as performed in Buenrostro et al. (2015) (~7–9 cycles). Post-PCR libraries were checked for proper library size distribution and concentration on TapeStation HS D1000, pooled to equimolar ratios, and SPRI purified and left-side size selected by 0.85× volume SPRI beads (>200 bp). Samples were sequenced paired-end on NovaSeq 6000.

Single-cell TIP-seq

TIP-seq for individual single cells was performed as in bulk TIP-seq, except binding of conA beads was omitted to permit FACS sorting of cells. For solution changes, centrifugation (500 x*g* for 3 min in a swing-bucket rotor) was used in place of magnetic bead separation. Briefly, cells were trypsinized, and a single-cell suspension was harvested and washed once with an equal volume of room-temperature PBS, then moved to a 2-ml tube and washed twice with room-temperature wash buffer. Cells were counted, and 100,000 cells were moved to a 1.5-ml LoBind Eppendorf tube and incubated with primary antibody (1:100) in antibody buffer overnight with rotation, incubated with secondary antibody (1:100) in Dig-wash buffer for 1 h at room temperature with rotation, then washed twice with 1 ml Dig-wash buffer and bound with pA-Tn5 (1:100) containing T7 promoter in Dig-300 buffer at room temperature with rotation. Cells were washed three times with 1 ml Dig-300 buffer to remove unbound pA-Tn5, and transposition was activated by adding 100 µl Tag buffer and incubated at 37°C for 1 h. Tagmentation was halted with the addition of 4 µl of 0.5 M EDTA (20 mM final), and cells were washed and resuspended in FACS buffer on ice. Single cells were FACS sorted into individual wells containing 10 µl PBS (+ EDTA) and underwent the remainder of TIP-seq as previously described in bulk.

sciTIP-seq

The majority of sciTIP-seq was performed as in bulk TIP-seq, except binding of conA beads was omitted to permit FACS sorting of cells. For solution changes, centrifugation (500 ×*g* for 3 min in a swing-bucket rotor) was used in place of magnetic bead separation. Briefly, cells were trypsinized, and a single-cell suspension was harvested and washed once with an equal volume of room-temperature PBS, then moved to a 2-ml tube and washed twice with room-temperature wash buffer. Approximately 100,000 cells were moved to a 1.5-ml LoBind Eppendorf tube and incubated with primary antibody (1:100) in antibody buffer overnight with rotation, then incubated with secondary antibody (1:100) in Dig-wash buffer for 30–60 min at room temperature with rotation. pA-Tn5–adapter complexes were prepared as in the pA-Tn5–adapter complex formation section. After secondary antibody binding, cells were washed twice with 1 ml Dig-wash buffer before distributing ~2,000 cells/well in a

96-well plate containing barcoded pA-Tn5–adapter complexes (up to four 96-well plates can be processed at a time with 384 uniquely barcoded pA-Tn5 transposons) in a final volume of 50 µl in Dig-300 buffer. Cells were incubated at room temperature for 1 h with rotation to allow pA-Tn5 to bind to antibody targets. Cells were washed three times for 5 min with 180 µl Dig-300 buffer to remove unbound pA-Tn5. Tagmentation was activated by resuspending cells in 20 µl Tag buffer and incubating for 1 h at 37°C. Tagmentation was terminated with the addition of 1 µl of 0.5 M EDTA (20 mM final concentration) and gentle agitation and incubated at room temperature for 15 min. Cells were gently pipetted up and down three times to mix and declump cells and then pooled in a single 2-ml LoBind Eppendorf tube. The wells of the 96-well plate were rinsed with 1× FACS buffer (2 mM EDTA, 1% BSA in PBS) to collect any residual cells and combined into the same 2-ml tube. Cells were either stored temporarily at −20°C or continued immediately for FACS redistribution. Cells were pelleted and resuspended in 1 ml of 1× FACS buffer containing 50 µl propidium iodide, then strained though 30-µm mesh to remove cell clumps. Cells were stored on ice until FACS sorting of 25–100 cells into each well of a 96-well plate containing 10 µl PBS. (The number of cells that can be pooled between indexing steps is a function of the number of possible index combinations used. As the number of possible index combinations increase, the theoretical probability of collisions decreases exponentially. Using more index 1 combinations decreases the probability of two cells receiving the same combination of barcodes (i.e., collision) during both indexing steps, thus increasing the number of cells that can be processed per well for index 2 addition. If only using 96 barcodes, 25 cells per well are sorted.)

DNA purification, IVT, cDNA synthesis, and fragmentation were all performed as described in bulk TIP-seq, except a different primer was used during second-strand synthesis (sss_sciTv2; Table S1) for compatibility with custom sciTIP transposons. Index 2 was added by PCR using 8 × 12 Nextera indexes. PCR reactions were prepared by adding 20 µl NEBNext High-Fidelity 2X PCR Master Mix, 2 µl of 10 µM custom scT7_S5XX index primers, and 2 µl of 10 µM standard i7 Nextera indexes, for a total volume of 40 µl. The optimal number of PCR cycles was determined through quantitative PCR as performed in Buenrostro et al. (2015) (typically 7–12 cycles). Post-PCR libraries were checked for proper library size distribution and concentration on TapeStation HS D1000. Libraries were pooled to equimolar ratios then purified and left-side size selected using SPRI beads (0.85× volume; >200 bp). Note for sequencing: The resulting library fragment structure is such that the P7 end contains a standard 8-bp i7 index, while the P5 end contains both a 6-bp r5 index (1–6 bp) and an 8-bp i5 index (21–29 bp), separated by a 15-bp spacer. Thus, the sciTIP-seq library pool was paired-end sequenced on a NovaSeq 6000 SP v1.5 flow cell using a modified run setup (see Table S2) to capture all necessary index sequence information for demultiplexing (51 cycles for read 1; 8 cycles for index 1 [i7]; 29 cycles for index 2 [i5 and r5]; 51 cycles for read 2).

TIP-seq/CUT&Tag data processing

TIP-seq and CUT&Tag data processing were performed as described in Kaya-Okur et al. (2019). Briefly, paired-end sequencing reads were aligned to hg38 or mm10 reference genomes using Bowtie 2 (Langmead and Salzberg, 2012) and filtered for PCR/optical duplicates using SAMtools markdup and fixmate (Danecek et al., 2021), and TIP-seq samples were filtered for T7 duplicate reads based on read 1 start positions using a custom script. BAM files were converted to BED files containing read positions and either converted to BigWig files for visualization in Integrative Genomics Viewer (IGV) or processed in parallel with sciTIP-seq and/or bulk ChIP-seq data.

sciTIP-seq data processing

Index sequences were appended to the header of their corresponding reads in read 1 and read 2 FASTQ files using bcl2fastq (see Table S2 for instructions on how to set up bcl2fastq to output FASTQ files with index sequences appended to read headers). Demultiplexing of single cells, read mapping and filtering, and subsequent data analysis were performed using custom R scripts. Read 1 and read 2 FASTQ files were demultiplexed to single-cell FASTQ files based on their unique combinations of i5, i7, and r5 index barcodes using a custom demultiplexing script. Sample subtypes were assigned according to which r5 barcodes were used on that subsample, and single-cell FASTQ files were renamed accordingly (i.e., subsample_S5XX_N7XX_r5XXX_R1/2.fastq). Single-cell paired-end FASTQ files were aligned to either mm10 or hg38 reference genomes using Bowtie 2 (Langmead and Salzberg, 2012), and reads were filtered for low-quality mapping scores and PCR and optical duplicates using SAMtools markdup and fixmate (Danecek et al., 2021). T7 duplicate reads were filtered based on read 1 start positions using a custom script. Cells containing <1,000 reads were removed for subsequent analysis. Pseudo-bulk samples (aggregate of all single cells) were created in R using bedCat from the R package travis (https://github.com/dvera/travis). For visualization in IGV, BED files were binned into 2.5-kb genomic windows using BEDTools coverage (Quinlan and Hall, 2010) and converted to BigWig files using bedgraphToBigWig. Peak calling was performed using MACS2 (Zhang et al., 2008) or SEACR (Meers et al., 2019), and peak overlap was plotted using R package ChIPpeakAnno (Zhu et al., 2010). Motif searching was conducted using MEME Suite (Bailey et al., 2009) on repeat-masked peak FASTA files. Enrichment heatmaps were made using BAMscale (Pongor et al., 2020) and deepTools computeMatrix (Ramírez et al., 2016), with peaks from ENCODE ChIP-seq as reference. FRiP was performed using custom R scripts to calculate the intersection of reads from each single cell with the pseudo-bulk peak files. For evaluation of doublet contamination (collision rate), single-cell FASTQ files from a mixed-species experiment were mapped to both mm10 and hg38 reference genomes using Bowtie 2, and cells with <200 reads were filtered out. Cells with >10% reads mapping to both reference genomes were classified as collisions. Allele parsing was performed using SNPsplit (Krueger and Andrews, 2016; https://www.bioinformatics.babraham.ac.uk/projects/SNPsplit) with the SNP VCF file downloaded from https://www.sanger.ac.uk/science/data/mouse-genomes-project. For a direct data comparison between sciTIP-seq and Paired-Tag, datasets were downloaded from the National Center for Biotechnology Information (NCBI) Gene

Expression Omnibus (GEO) under accession number GSE152020 and preprocessed as described in Zhu et al. (2010) using custom scripts from the study (available at https://github.com/cxzhu/Paired-Tag). Demultiplexed single cells for H3K27me3 and H3K9me3 from Paired-Tag sublibrary 1 were processed and analyzed in parallel with sciTIP-seq single cells using identical filters as the Paired-Tag manuscript and identical peak calling parameters between each library type. scRNA-seq data were downloaded from GEO under accession no. GSM3160745 and processed using Cell Ranger and Seurat (Hao et al., 2021) with RNAPII sciTIP-seq data to produce UMAPs. scRNA-seq cells were randomly downsampled to match the number of RNAPII sciTIP-seq cells.

Antibodies

Antibodies used in this study were rabbit anti-H3K27me3 monoclonal antibody (mAb; 9733; Cell Signaling Technology); rabbit anti-H3K9me3 polyclonal antibody (pAb; ab8898; Abcam); rabbit anti-CTCF pAb (07-729; Millipore); rabbit anti-RNAPII-Ser2Ph mAb (13499; Cell Signaling Technology); rabbit anti-H3K27ac pAb (ab4729; Abcam); and guinea pig anti-rabbit IgG pAb (ABIN101961; antibodies-online).

Online supplemental materials

Fig. S1 shows Pearson correlation matrices and Venn diagrams of peak overlap among bulk TIP-seq samples and the MEME logos from motif searches of CTCF samples. Fig. S2 shows various sciTIP-seq quality metrics, including unique read counts before filtering out low-read cells, percent unique reads before and after filtering of low-read cells, PCR and T7 duplication rate vs. read count of single cells, and barnyard analysis of collision rates. Fig. S3 shows a comparison of unique reads counts and FRiPs between sciTIP-seq and Paired-Tag datasets and UMAPs of RNAPII sciTIP-seq vs. scRNA-seq. Table S1 shows a list of all oligonucleotides used in this study. Table S2 shows bcl2fastq instructions on how to append index sequence to FASTQ read headers for demultiplexing.

Data availability

The sequencing data obtained in this study have been deposited at the NCBI GEO under accession number GSE188512. ENCODE (https://www.encodeproject.org/) ChIP-seq datasets were downloaded with the following accession numbers: HCT116 CTCF (ENCBS409ENC) and mESC H3K9me3 (ENCSR000CFZ). 4D Nucleome (https://www.4dnucleome.org/) ChIP-seq datasets were downloaded with the following accession numbers: HCT116 H3K27me3 (4DNESBI57YML), mESC H3K27me3 (4DNESKIB4QKT), mESC RNAPII (4DNESW1G42GW), mESC H3K27ac (4DNESEUGAD5F), and mESC CTCF (4DNESD1LH7J9). Other external datasets were downloaded from NCBI GEO with the following accession numbers: Paired-Tag (GSE152020) and scRNA-seq (GSE114952). Scripts for single-cell demultiplexing and data analysis are available at https://github.com/dbart1807/TIP-seq.

Acknowledgments

We thank Kami Ahmad for early guidance and valuable comments on data, Jesse Turner for cell culture assistance, Takayo Sasaki for assistance with sequence library pooling and handling, Steven Miller and Amber Brown for assistance with MiSeq pilot runs and Nextera indexes, Cynthia Vied and Yanming Yang for help with Illumina NovaSeq 6000 and bcl2fastq troubleshooting, and Ruth Didier and Beth Alexander for flow cytometry assistance.

This work was funded by National Institutes of Health grant R21 HG010403 (D.M. Gilbert), Japan Science and Technology Agency Core Research for Evolutional Science and Technology grant JPMJCR16G1 (Y. Ohkawa and H. Kimura), and Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research grant JP18H05527 (Y. Ohkawa and H. Kimura).

The authors declare no competing financial interests.

Author contributions: D.A. Bartlett and D.M. Gilbert conceived the experiments; D.A. Bartlett designed and executed all experiments and analyses; D.A. Bartlett and D.M. Gilbert wrote the manuscript; V. Dileep wrote the cell demultiplexing script; S. Henikoff provided pA-Tn5, valuable consultation, and review and editing of the manuscript; and H. Kimura, T. Handa, Y. Ohkawa provided ChIL-seq training that inspired the design and implementation of TIP-seq.

Submitted: 13 March 2021
Revised: 15 September 2021
Accepted: 27 October 2021

References

Bailey, T.L., M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, and W.S. Noble. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37:W202–W208. https://doi.org/10.1093/nar/gkp335

Baranello, L., F. Kouzine, S. Sanford, and D. Levens. (2016). ChIP bias as a function of cross-linking time. *Chromosome Res.* 24:175–181. https://doi.org/10.1007/s10577-015-9509-1

Bartosovic, M., M. Kabbe, and G. Castelo-Branco. 2021. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* 39:825–835. https://doi.org/10.1038/s41587-021-00869-9

Brind'Amour, J., S. Liu, M. Hudson, C. Chen, M.M. Karimi, and M.C. Lorincz. 2015. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* 6:6033. https://doi.org/10.1038/ncomms7033

Cao, Z., C. Chen, B. He, K. Tan, and C. Lu. 2015. A microfluidic device for epigenomic profiling using 100 cells. *Nat. Methods.* 12:959–962. https://doi.org/10.1038/nmeth.3488

Buenrostro, J.D., B. Wu, H.Y. Chang, and W.J. Greenleaf. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol.* 109:21.29.1-21.29.9. . https://doi.org/10.1002/0471142727.mb2129s109

Cao, J., J.S. Packer, V. Ramani, D.A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S.N. Furlan, F.J. Steemers, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science.* 357:661–667. https://doi.org/10.1126/science.aam8940

Carter, B., W.L. Ku, J.Y. Kang, G. Hu, J. Perrie, Q. Tang, and K. Zhao. 2020. Author correction: mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). *Nat. Commun.* 11:4424.)https://doi.org/10.1038/s41467-019-11559-1

Chen, C., D. Xing, L. Tan, H. Li, G. Zhou, L. Huang, and X.S. Xie. 2017. Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* 356:189–194.

Cusanovich, D.A., R. Daza, A. Adey, H.A. Pliner, L. Christiansen, K.L. Gunderson, F.J. Steemers, C. Trapnell, and J. Shendure. 2015. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science.* 348:910–914. https://doi.org/10.1126/science.aab1601

Danecek, P., J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, R.M. Davies, and H. Li. 2021.

Twelve years of SAMtools and BCFtools. *Gigascience*. 10:giab008. https://doi.org/10.1093/gigascience/giab008

Eberwine, J., H. Yeh, K. Miyashiro, Y. Cao, S. Nair, R. Finnell, M. Zettel, and P. Coleman. 1992. Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA*. 89:3010–3014. https://doi.org/10.1073/pnas.89.7.3010

Gopalan, S., Y. Wang, N.W. Harper, M. Garber, and T.G. Fazzio. 2021. Simultaneous profiling of multiple chromatin proteins in the same cells. *bioRxiv*. (Preprint posted April 28, 2021) https://doi.org/10.1101/2021.04.27.441642

Grosselin, K., A. Durand, J. Marsolier, A. Poitou, E. Marangoni, F. Nemati, A. Dahmani, S. Lameiras, F. Reyal, O. Frenoy, et al. 2019. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* 51:1060–1066. https://doi.org/10.1038/s41588-019-0424-9

Hainer, S.J., A. Bošković, K.N. McCannell, O.J. Rando, and T.G. Fazzio. 2019. Profiling of Pluripotency factors in single cells and early embryos. *Cell*. 177:1319–1329.e11. https://doi.org/10.1016/j.cell.2019.03.014

Hao, Y., S. Hao, E. Andersen-Nissen, W.M. Mauck III, S. Zheng, A. Butler, M.J. Lee, A.J. Wilk, C. Darby, M. Zager, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell*. 184:3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Harada, A., K. Maehara, T. Handa, Y. Arimura, J. Nogami, Y. Hayashi-Takanaka, K. Shirahige, H. Kurumizaka, H. Kimura, and Y. Ohkawa. 2019. A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.* 21:287–296. https://doi.org/10.1038/s41556-018-0248-3

Hashimshony, T., F. Wagner, N. Sher, and I. Yanai. 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2:666–673. https://doi.org/10.1016/j.celrep.2012.08.003

Hoeijmakers, W.A.M., R. Bártfai, K.J. Françoijs, and H.G. Stunnenberg. 2011. Linear amplification for deep sequencing. *Nat. Protoc.* 6:1026–1036. https://doi.org/10.1038/nprot.2011.345

Kaya-Okur, H.S., S.J. Wu, C.A. Codomo, E.S. Pledger, T.D. Bryson, J.G. Henikoff, K. Ahmad, and S. Henikoff. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* 10:1930. https://doi.org/10.1038/s41467-019-09982-5

Krueger, F., and S.R. Andrews. 2016. SNPsplit: allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000 Res.* 5:1479. https://doi.org/10.12688/f1000research.9037.1

Ku, W.L., K. Nakamura, W. Gao, K. Cui, G. Hu, Q. Tang, B. Ni, and K. Zhao. 2019. Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat. Methods*. 16:323–325. https://doi.org/10.1038/s41592-019-0361-7

Ku, W.L., L. Pan, Y. Cao, W. Gao, and K. Zhao. 2021. Profiling single-cell histone modifications using indexing chromatin immunocleavage sequencing. *Genome Res.* 31:1831–1842. https://doi.org/10.1101/gr.260893.120

Lake, B.B., S. Chen, B.C. Sos, J. Fan, G.E. Kaeser, Y.C. Yung, T.E. Duong, D. Gao, J. Chun, P.V. Kharchenko, and K. Zhang. 2018. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* 36:70–80. https://doi.org/10.1038/nbt.4038

Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. 9:357–359. https://doi.org/10.1038/nmeth.1923

Lareau, C.A., F.M. Duarte, J.G. Chew, V.K. Kartha, Z.D. Burkett, A.S. Kohlway, D. Pokholok, M.J. Aryee, F.J. Steemers, R. Lebofsky, and J.D. Buenrostro. 2019. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* 37:916–924. https://doi.org/10.1038/s41587-019-0147-6

Marinov, G.K. 2018. A decade of ChIP-seq. *Brief. Funct. Genomics*. 17:77–79. https://doi.org/10.1093/bfgp/ely012

Meers, M.P., D. Tenenbaum, and S. Henikoff. 2019. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin*. 12:42. https://doi.org/10.1186/s13072-019-0287-4

Meers, M.P., D.H. Janssens, and S. Henikoff. 2021. Multifactorial chromatin regulatory landscapes at single cell resolution. *bioRxiv*. (Preprint posted July 9, 2021) https://doi.org/10.1101/2021.07.08.451691

Minkina, A., and J. Shendure. 2019. Expanding the single-cell genomics toolkit. *Nat. Genet.* 51:931–932. https://doi.org/10.1038/s41588-019-0429-4

Mulqueen, R.M., D. Pokholok, S.J. Norberg, K.A. Torkenczy, A.J. Fields, D. Sun, J.R. Sinnamon, J. Shendure, C. Trapnell, B.J. O'Roak, et al. 2018. Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* 36:428–431. https://doi.org/10.1038/nbt.4112

Pongor, L.S., J.M. Gross, R. Vera Alvarez, J. Murai, S.-M. Jang, H. Zhang, C. Redon, H. Fu, S.-Y. Huang, B. Thakur, et al. 2020. BAMscale: quantification of next-generation sequencing peaks and generation of scaled coverage tracks. *Epigenetics Chromatin*. 13:21. https://doi.org/10.1186/s13072-020-00343-x

Quinlan, A.R., and I.M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26:841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramani, V., X. Deng, R. Qiu, K.L. Gunderson, F.J. Steemers, C.M. Disteche, W.S. Noble, Z. Duan, and J. Shendure. 2017. Massively multiplex single-cell Hi-C. *Nat. Methods*. 14:263–266. https://doi.org/10.1038/nmeth.4155

Ramírez, F., D.P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dündar, and T. Manke. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44:W160–W165. https://doi.org/10.1093/nar/gkw257

Rooijers, K., C.M. Markodimitraki, F.J. Rang, S.S. De Vries, A. Chialastri, K.L. De Luca, D. Mooijman, S.S. Dey, and J. Kind. 2019. Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat. Biol.* 37:766–772. https://doi.org/10.1038/s41587-019-0150-y

Skene, P.J., and S. Henikoff. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*. 6:21856. https://doi.org/10.7554/eLife.21856

Skene, P.J., J.G. Henikoff, and S. Henikoff. 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* 13:1006–1019. https://doi.org/10.1038/nprot.2018.015

Sos, B.C., H.L. Fung, D.R. Gao, T.F. Osothprarop, A. Kia, M.M. He, and K. Zhang. 2016. Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. *Genome Biol.* 17:20. https://doi.org/10.1186/s13059-016-0882-7

Tang, G.Q., R.P. Bandwar, and S.S. Patel. 2005. Extended upstream A-T sequence increases T7 promoter strength. *J. Biol. Chem.* 280:40707–40713. https://doi.org/10.1074/jbc.M508013200

van Bakel, H., F.J. van Werven, M. Radonjic, M.O. Brok, D. van Leenen, F.C.P. Holstege, and H.T.M. Timmers. 2008. Improved genome-wide localization by ChIP-chip using double-round T7 RNA polymerase-based amplification. *Nucleic Acids Res.* 36:e21. https://doi.org/10.1093/nar/gkm1144

van Galen, P., A.D. Viny, O. Ram, R.J.H. Ryan, M.J. Cotton, L. Donohue, C. Sievers, Y. Drier, B.B. Liau, S.M. Gillespie, et al. 2016. A multiplexed system for quantitative comparisons of chromatin landscapes. *Mol. Cell.* 61:170–180. https://doi.org/10.1016/j.molcel.2015.11.003

Van Gelder, R.N., M.E. von Zastrow, A. Yool, W.C. Dement, J.D. Barchas, and J.H. Eberwine. 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA*. 87:1663–1667. https://doi.org/10.1073/pnas.87.5.1663

Vitak, S.A., K.A. Torkenczy, J.L. Rosenkrantz, A.J. Fields, L. Christiansen, M.H. Wong, L. Carbone, F.J. Steemers, and A. Adey. 2017. Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat. Methods*. 14:302–308. https://doi.org/10.1038/nmeth.4154

Wang, Q., H. Xiong, S. Ai, X. Yu, Y. Liu, J. Zhang, and A. He. 2019. CoBATCH for high-throughput single-cell epigenomic profiling. *Mol. Cell.* 76:206–216.e7. https://doi.org/10.1016/j.molcel.2019.07.015

Weiner, A., D. Lara-Astiaso, V. Krupalnik, O. Gafni, E. David, D.R. Winter, J.H. Hanna, and I. Amit. 2016. Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. *Nat. Biotechnol.* 34:953–961. https://doi.org/10.1038/nbt.3652

Wu, S.J., S.N. Furlan, A.B. Mihalas, H.S. Kaya-Okur, A.H. Feroze, S.N. Emerson, Y. Zheng, K. Carson, P.J. Cimino, C.D. Keene, et al. 2021. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat. Biotechnol.* 39:819–824. https://doi.org/10.1038/s41587-021-00865-z

Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, and X.S. Liu. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang, B., H. Zheng, B. Huang, W. Li, Y. Xiang, X. Peng, J. Ming, X. Wu, Y. Zhang, Q. Xu, et al. 2016. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*. 537:553–557. https://doi.org/10.1038/nature19361

Zhu, L.J., C. Gazin, N.D. Lawson, H. Pagès, S.M. Lin, D.S. Lapointe, and M.R. Green. 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*. 11:237. https://doi.org/10.1186/1471-2105-11-237

Zhu, C., Y. Zhang, Y.E. Li, J. Lucero, M.M. Behrens, and B. Ren. 2021. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat. Methods*. 18:283–292. https://doi.org/10.1038/s41592-021-01060-3

Supplemental material

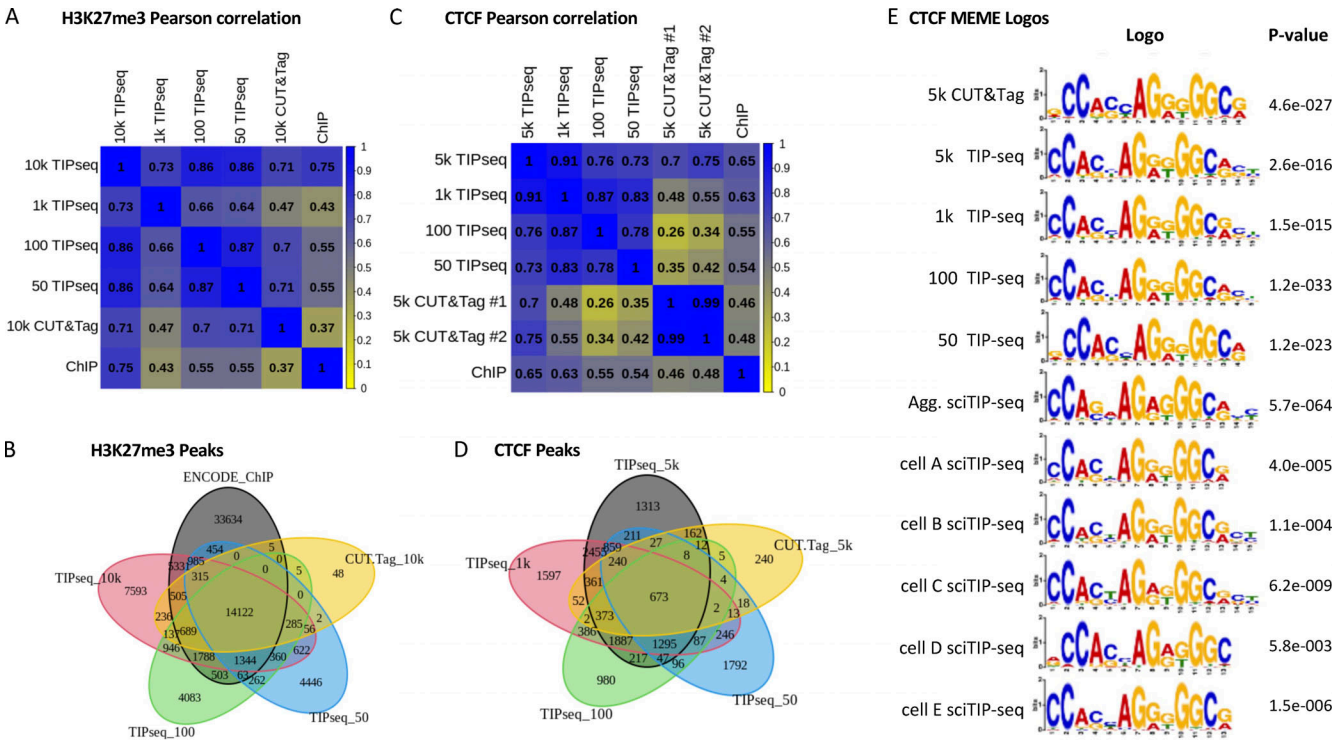
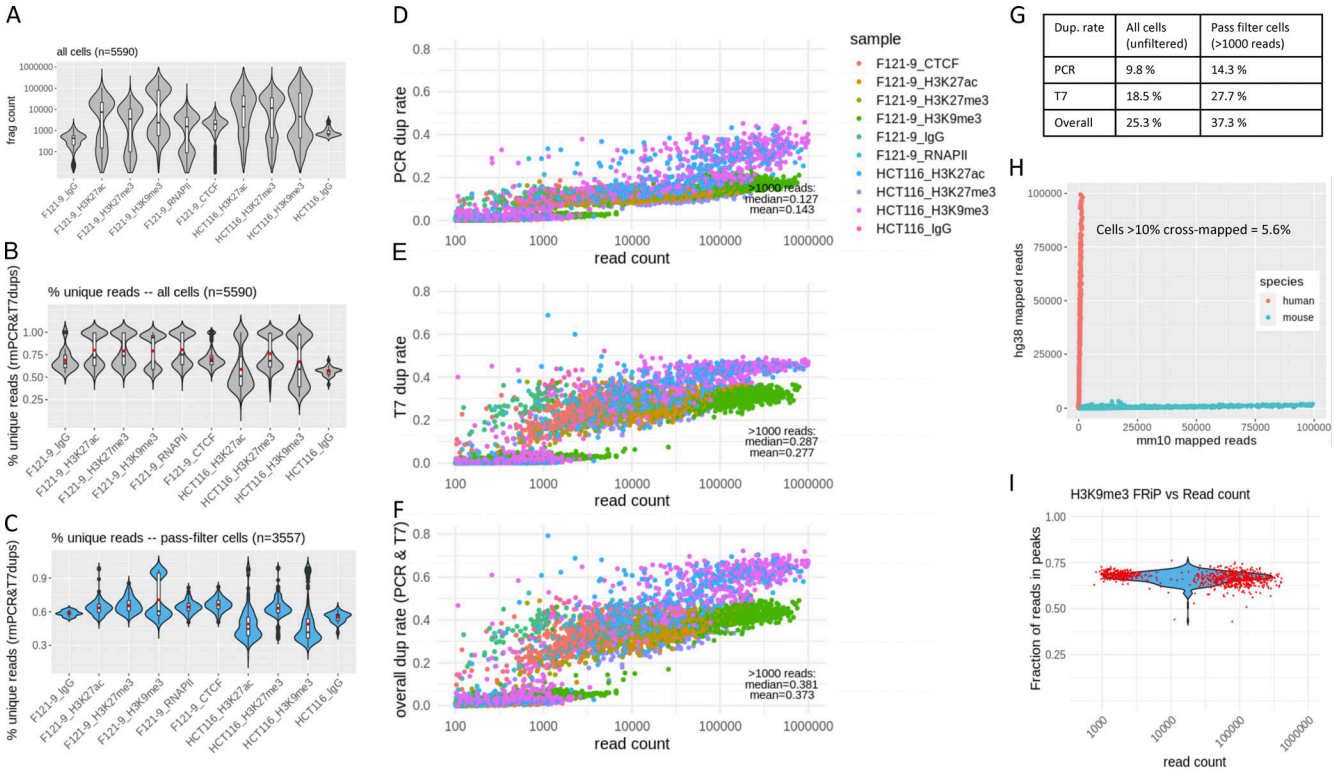


Figure S1. **Bulk TIP-seq quality metrics. (A)** H3K27me3 Pearson correlations among TIP-seq, CUT&Tag, and ENCODE ChIP-seq targeting H3K27me3 in an HCT116 cell using 50-kb bins. **(B)** H3K27me3 Venn diagram displaying peak overlap among samples. **(C)** CTCF Pearson correlations among TIP-seq, CUT&Tag, and ENCODE ChIP-seq targeting TF CTCF in an HCT116 cell using 50-kb bins. **(D)** CTCF Venn diagram displaying peak overlap among samples. **(E)** MEME logo representations for CTCF bulk, single-cell aggregate (Agg.), and five single cells chosen at random showing the most prevalent motif identified for each respective library and the P value associated with it.



Dup. rate	All cells (unfiltered)	Pass filter cells (>1000 reads)
PCR	9.8 %	14.3 %
T7	18.5 %	27.7 %
Overall	25.3 %	37.3 %

Figure S2. **sciTIP-seq quality metrics. (A)** Violin plots showing the number of unique reads per cell for each sample before filtering out low-read cells. **(B)** Violin plots showing the percent unique reads for each sample after filtering out low-read cells. **(C)** Violin plots showing the percent unique reads for each sample after filtering out low-read cells. **(D)** PCR duplication rate vs. read count for each single cell (samples grouped by color). Mean and median duplication rates (after filtering out cells with <1,000 reads). **(E)** T7 duplication rate vs. read count for each single cell (samples grouped by color). **(F)** Overall (PCR and T7) duplication rate vs. read count for each single cell (samples grouped by color). Mean and median duplication rates (after filtering out cells with <1,000 reads) are shown in bottom right corner of duplication scatterplots D–F. **(G)** Table displaying the mean PCR, T7, and overall duplication rates before and after filtering of low-reads cells. **(H)** Barnyard analysis scatterplot displaying cross mapping of mouse and human cells to mm10 and hg38 reference genomes. 5.6% of cells had >10% reads mapping to both reference genomes after removal of cells with <200 reads. **(I)** FRiP violin plot with cell read counts (x axis) showing that the cohort of cells with fewer reads retains a high signal-to-noise ratio. dup, duplication; frag, fragment.

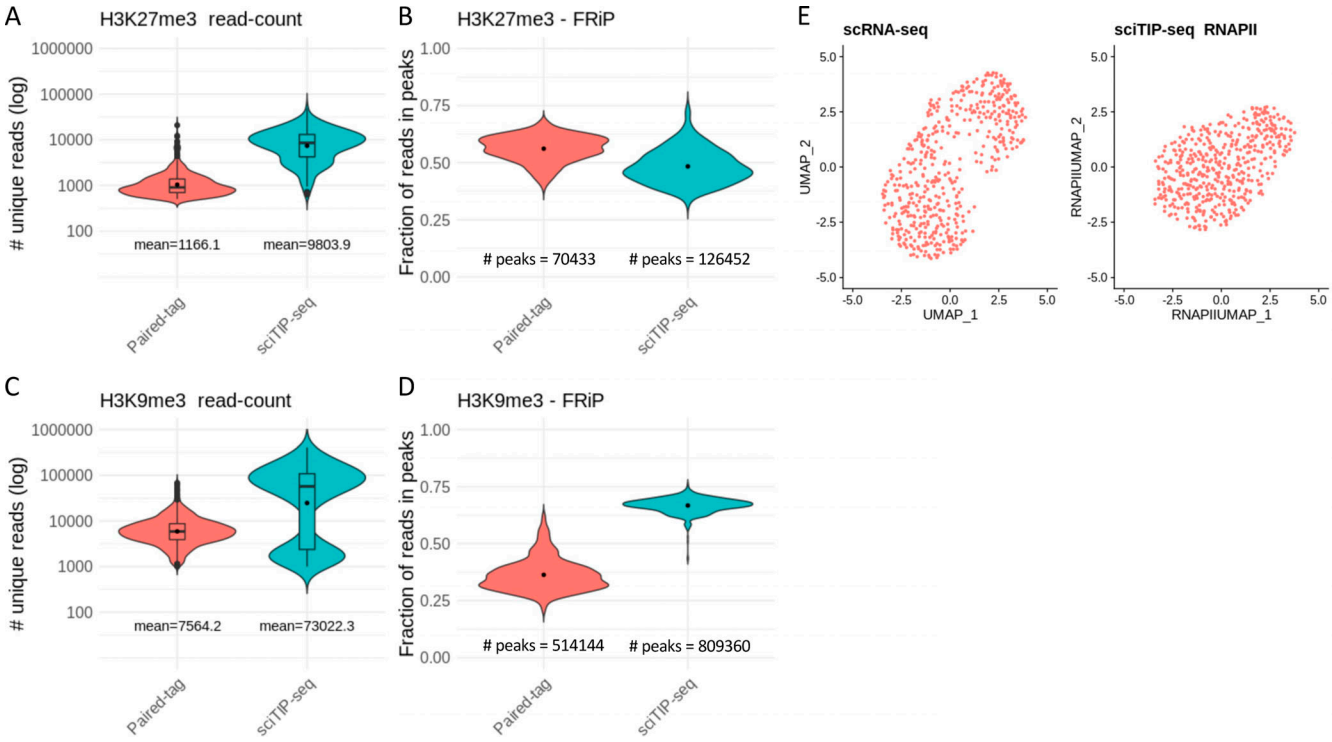


Figure S3. **Comparison to Paired-Tag and scRNA-seq.** Library comparisons between downloaded Paired-Tag from mouse hippocampus cells (red) and TIP-seq from F121-9 mESCs (blue) using identical filters and peak calling parameters between both library types. **(A and B)** Violin plots showing unique read counts and FRiP for H3K27me3. **(C and D)** Violin plots showing unique read counts and FRiP for H3K9me3. Means are marked with a black dot. Peak numbers provided below FRiP violins. **(E)** UMAPs of 10x Genomics scRNA-seq data (GEO accession no. GSM3160745) and RNAPII sciTIP-seq data. scRNA-seq cell numbers were randomly downsampled to match that of RNAPII sciTIP-seq (480 cells).

Table S1, provided online as a separate Excel file, shows a list of all oligonucleotides used in this study. Table S2, provided online as a separate Word file, shows instructions for appending index sequences to FASTQ read headers to enable demultiplexing of single cells.